

Djamel Edine YAGOUBI  
Campus Saint-Priest – Bâtiment 5  
860 rue de St Priest  
34392 Montpellier Cedex 5, France

☎ +33 7 58 68 49 91

✉ yagoubi.djamel@gmail.com

💻 djameledine-yagoubi.info

🌐 www.linkedin.com/in/ydjameledine/

🔗 https://github.com/djamelinfo

**Data Scientist & Data Engineer**  
Traitement et analyse de big data

## Déclaration Personnelle

Bientôt titulaire d'un doctorat en informatique (dans le domaine du big data). Passionné par les nouvelles technologies et leur mise en œuvre pour traiter des problèmes difficiles. Créative, notamment dans l'utilisation des techniques en statistique, data mining et gestion de données pour produire des solutions innovantes. Grande expérience dans le domaine du big data (Hadoop, Spark).

Mon objectif est de conforter mon expérience de data scientist dans des domaines d'applications riches en problèmes stimulants.

## Expériences (Projets Réalisés)

Depuis octobre 2014 : **Doctorant, Equipe Zenith, Inria & LIRMM, Montpellier**

L'équipe Zenith (<https://team.inria.fr/zenith>) est internationalement reconnue dans la gestion, le traitement et l'analyse de big data, avec de fortes collaborations industrielles.

**Contexte :** le projet européen CloudDBAppliance (European Cloud In-Memory Database) a démarré fin 2016 dans le cadre du programme H2020 (<http://cloudadb.eu/>, <https://www.linkedin.com/company/cloudbappliance/>) avec les entreprises Bull, LeanXcale, ActiveViam, BTO, Wind Tre S.p.A, IKEA, INESC TEC, Universidad Politécnica de Madrid, INRIA, QIVOS, Singular Logic.

Projet 1 : ***Indexation massivement distribuée de séries temporelles basée sur l'index iSAX***

<http://djameledine-yagoubi.info/projects/DPiSAX/>

- **Motivation :** Les séries temporelles ont un rôle important dans la résolution de plusieurs problèmes du monde réel. Par exemple les individus peuvent surveiller différents indicateurs pour leurs activités personnelles (via des compteurs intelligents ou des prises intelligentes pour la consommation d'électricité ou d'eau), ou des activités professionnelles (à travers les capteurs installés sur les plantes par les agriculteurs).
- **Problème :** Répondre à des requêtes de similitude sur des centaines de millions de séries temporelles est un défi qui nécessite à la fois des techniques d'indexation efficaces et du calcul parallèle.
- **Solution :** Conception et réalisation d'une approche basée sur l'index iSAX qui passe à l'échelle de manière presque linéaire dans des environnements parallèles et fournit des réponses de haute qualité.
- **Outils utilisés :** Java, Scala, Hadoop (HDFS, YARN), Spark Core, Grid 5K (<https://www.grid5000.fr>), Nef Cluster (<https://wiki.inria.fr/ClustersSophia>).

Projet 2 : ***Indexation massive de séries temporelles basée sur la projection aléatoire***

<https://radiussketch.github.io/RadiusSketch/>

- **Motivation :** Les séries temporelles se posent dans de nombreux domaines d'application tels que la finance, l'agronomie, la santé, la surveillance de la Terre, la prévision météorologique.
- **Problème :** En raison des progrès de la technologie des capteurs, les applications peuvent produire des millions à des trillions de séries temporelles par jour, ce qui nécessite des techniques d'indexation

et d'analyse parallèles très rapides.

- **Solution :** La réalisation d'une approche basée sur les "sketches" et la projection aléatoire mise en évidence par un passage à l'échelle de manière presque linéaire dans des environnements hautement distribués.
- **Outils utilisés :** Java, Scala, Hadoop (HDFS, YARN), Spark Core, Grid 5K (<https://www.grid5000.fr>), Nef Cluster (<https://wiki.inria.fr/ClustersSophia>).

#### Projet 3 : *Identification des paires de séries temporelles similaires dans un flux continu de données*

<http://djameledine-yagoubi.info/projects/ParCorr/>

- **Motivation :** Trouver des paires de séries temporelles similaires dans un flux continu de données est utile pour de nombreuses applications. Par exemple la fusion de capteurs, la négociation financière ou la surveillance de réseaux de communication.
- **Problème :** La détection de corrélation en parallèle dans un flux continu de séries temporelles est un défi qui nécessite des méthodes d'optimisation mathématique très fiables.
- **Solution :** Conception et réalisation d'un algorithme pour la détection parallèle de séries temporelles corrélées à partir des grilles distribuées en minimisant à la fois la taille et le nombre de messages nécessaires à la détection des candidats.
- **Outils utilisés :** Java, Scala, kafka, Spark Streaming, Grid 5K (<https://www.grid5000.fr>), Nef Cluster (<https://wiki.inria.fr/ClustersSophia>).

Mai 2014 - Juillet 2014 : **Stagiaire, Laboratoire d'informatique d'ORAN (LIO), Université d'Oran 1, Oran**

#### Projet : *Gestion de la consommation d'énergie dans les Clouds Computing*

- **Motivation :** Le Cloud computing est l'une des technologies modélisée à fournir un service plutôt qu'un produit, des services comme le calcul, le logiciel, l'accès aux données et le stockage. Les grands data centers virtualisés sont établis pour répondre à cette exigence.
- **Problème :** La demande croissante d'infrastructure dans les clouds computing augmente l'énergie consommée par les Data Centers, ce qui est devenu un problème critique. La forte consommation d'énergie se traduit par des coûts plus élevés des revenus et d'émissions de gaz CO<sub>2</sub>, cela résulte en augmentation de réchauffement de la planète dans un avenir proche.
- **Solution :** La réalisation d'un algorithme adoptant deux approches d'optimisation et de consommation d'énergie basées sur deux technologies : la virtualisation et la migration des machines virtuelles d'une machine physique vers une autre.
- **Outils utilisés :** Java, CloudSim 2.1.

Mai 2012 - Juin 2012 : **Stagiaire, Laboratoire d'informatique d'ORAN (LIO), Université d'Oran 1, Oran**

- Conception d'une nouvelle approche d'ordonnancement pour accélérer le traitement des paquets dans les routeurs, en utilisant plusieurs techniques d'optimisation simultanément.
- Outils utilisés : C++, TCL, OTCL, NS-2.35.

### Publication Scientifique

Novembre 2017	Djamel Edine Yagoubi, Reza Akbarinia, Florent Massegli, Themis Palpanas. <b>DPI-SAX : Massively Distributed Partitioned iSAX.</b> <i>IEEE International Conference on Data Mining (ICDM)</i>
Octobre 2017	Djamel Edine Yagoubi, Reza Akbarinia, Florent Massegli, Dennis Shasha. <b>RadiusSketch : Massively Distributed Indexing of Time Series.</b> <i>IEEE International Conference on Data Science and Advanced Analytics (DSAA)</i>
Septembre 2014	Djouhra Dad, Djamel Edine Yagoubi, Ghalem Belalem. <b>Energy Efficient VM Live Migration and Allocation at Cloud Data Centers.</b> <i>International Journal of Cloud Applications and Computing (IJCAC)</i>

## Education

**2014-2018**  
(en cours)

Inria (Equipe Zenith), Université De Montpellier. **Doctorat informatique.**  
Indexation et analyse de très grandes masses de séries temporelles.

**2012-2014**  
(2 ans)

Université d'Oran 1, Oran. **Master Informatique.**  
Théorie et Pratique des Grands Réseaux.  
▸ *Architecture Parallèle et Réseaux de Neurones.* ▸ *Paradigmes des Systèmes Répartis.* ▸ *Recherche Opérationnelle et Théorie Graphes.* ▸ *Théorie des systèmes répartis.* ▸ *Base de données Orientée Objet et Langage naturel.* ▸ *Systèmes d'Exploitation Avancés.* ▸ *La sécurité dans les réseaux.* ▸ *Intelligent Artificielle Distribuée.* ▸ *Manipulation et Indexation de données réparties.* ▸ *Cryptage des données.*

**2009-2012**  
(3 ans)

Université d'Oran 1, Oran. **Licence Mathématique Informatique.**  
Spécialisation en informatique.  
▸ *Programmation système.* ▸ *Conception et programmation objet.* ▸ *Théorie des langages.* ▸ *Théorie des graphes.* ▸ *Langages interprétés.* ▸ *Programmation logique.* ▸ *Algorithmique.* ▸ *Algorithmique avancée et complexité.* ▸ *Programmation linéaire.* ▸ *Programmation des interfaces graphiques.* ▸ *SGBD.* ▸ *Théorie des langages.* ▸ *Logique Mathématique.* ▸ *Probabilité et Statistique.*

**2009**

Lycée de Oued Tlelat, Oran. **Baccalauréat série science.**

## Compétences

*Big Data*

Hadoop, MapReduce, Spark Core, Spark ML, Spark Streaming, Kafka, Apache Mesos, Grid 5K, Nef Cluster, OAR task manager.

*Langages*

Java, Scala, Python, C, C++ (Qt), C#, Shell, R, TCL/OTCL, HTML/CSS, JavaScript, Matlab.

*Statistiques*

R.

*Office*

LaTeX, Markdown(Mou), LibreOffice, Microsoft Office.

*Bases de données*

Relationnelles (MySQL, SQLite, PostgreSQL, SQLServer), NoSQL (MongoDB, InfluxDB)

*Systèmes*

Mac OS, Unix, Linux (Ubuntu, Debian, Fedora), Windows.

*Outils liés au développement*

Git, SVN, Eclipse, Netbeans, IntelliJ Idea, Atom, Sublime Text 2/3, QtCreator.

*Analyse*

UML.

*Autres*

Adobe Photoshop, Adobe Illustrator.

## Langues et divers

*Anglais*

Courrant.

*Arabe*

Courrant.

*centres d'intérêt*

Bricolage, Mécanique Automobile, Open Source, Programmation